

A MODEL FOR GRADING BAR EXAMINATION ESSAY QUESTIONS

by Jean C. Gaskill

A number of articles appearing in *The Bar Examiner* have discussed the relative merits of the "analytical" versus the "holistic" methods of grading bar examination essay questions.¹ The former method utilizes a checklist approach in which the graders assign a certain weight to each of the issues the examinees are expected to cover, scout the answer for the presence or absence of those issues, add up the points and assign a grade depending upon the cumulative points awarded. The holistic method, on the other hand, utilizes a less particularized approach, looking to the overall content and quality of the answer.²

Research has shown that the analytical and holistic methods are equally reliable.³ Both methods require the graders to be thoroughly conversant with the issues the author intended to cover and how the author wanted them resolved. Both methods measure the examinee's ability to recognize and deal with the issues raised in the question, and both methods inherently compare how well one examinee did relative to others.

Neither method can claim scientific precision. Graders utilizing either method will still differ on how well examinees dealt with a particular issue; e.g., whether, although an examinee recognized the issue, he or she covered it well enough to receive full or partial credit.

The difference is that the holistic method is significantly faster than the analytic method. It is often possible to do two holistic ratings in the time it takes to do one analytic rating.⁴ If graders can save time grading

examinations without sacrificing consistency and reliability, there is no reason not to use the holistic method. It requires a mind-set shift and some training, but experience has shown that, once the graders become comfortable with it, they prefer it, and it produces the same scoring results as the more tedious analytical method.

What follows is a description of a model for grading essay questions holistically. It is a variation of the highly successful method used at the semi-annual grading workshops attended by the graders from jurisdictions that administer the NCBE's Multistate Essay Examination and the method that will in all probability be used at the grading workshops for the upcoming Multistate Performance Test. It is adaptable to almost any essay grading medium.

PREPARATION

As has frequently been said about most successful ventures in life, there are three essential prerequisites: preparation, preparation and preparation. It is no less the case when it comes to grading essay examinations.

The Written Narrative Analysis

The goal of any grading procedure is, of course, to achieve fairness and consistency among the grades assigned, and, in order to ensure these qualities, it is imperative that the graders be thoroughly familiar with the question.

Before a particular question is even adopted for use on the bar examination, the author should have re-

BEST AVAILABLE COPY

searched the issues carefully and written a narrative analysis, supported by legal authorities, of all the issues he or she expects the examinees to discuss. If the author has done so, it can be assumed that he or she is already thoroughly familiar with the question and the analysis when it comes time to do the grading.

However, if the author of the question is not the person or the only person who will be doing the grading, something must be done to ensure that all the graders have immersed themselves in the question. One way is to give them the author's analysis and require them to study it before beginning the grading.

A better way, albeit one that requires more work, is to send the graders the question they are going to be grading right after the administration of the examination, unaccompanied by the author's analysis, and require them to conduct independent research and write their own answers to the question.

The graders should then exchange their independent analyses and confer among themselves to reconcile the inevitable differences that arise whenever two or more lawyers opine on the same problem. Their discussions should result in what is often called a "consensus analysis," the sort of "official" answer against which the examinees' answers will be graded.

Familiarization Papers

Before the actual grading begins, the graders should read a number of answers (perhaps five or ten) drawn at random, not to assign grades to them but just to see how the examinees are handling the question. These may or may not turn out to be representative of the rest of the papers, but they will give the graders a fairly good idea of whether there are issues some of the examinees have missed, whether there are issues they have picked up that the author and the graders did not, and whether there are latent ambiguities that the author and the graders failed to discern.

If the graders perceive any such anomalies in the course of reading the familiarization papers, they will probably want to adjust the grading standards to account for them. For example, if it turns out that most applicants are missing a point that the author thought was clearly raised in the question, it might be that it is more subtle than the author intended. In such a case, the graders might decide that it should be discounted in the grading or that it should be treated purely as an "extra credit" issue for the applicants who discuss it, without deducting credit from the answers of those who do not. Similarly, if a point that the author did not intend turns up regularly in the answers, the graders need to decide whether it is a valid point and whether to give credit for it.

This is not to say that the tail should wag the dog, but it is necessary for graders to be sensitive to the possibility that what the author thought was unambiguously raised is in reality not so. It is always a judgment call whether to discount or credit a latent issue in the grading process, but if fairness and consistency are to be maintained over the grading continuum, such problems have to be discovered and the decision made early on. One of the benefits of reading a number of familiarization papers is that it makes it possible to discover such problems before the grading begins and to adjust the grading standards to take them into account.¹

Grading Guidelines

At some point in the process, it is necessary to decide how much relative weight to give each of the issues in the question. Irrespective of which answers are being graded on an absolute scale (answer 2 passes, answer 1 fails) or on a relative scale (answer 2 is better than answer 1), it is desirable to commit to writing for the guidance of the graders a description of an acceptable answer, including the characteristics that separate a "bare pass" from an answer entitled to greater credit. For example, if the papers are being scored on a six-point scale and a score

¹ A Model for Grading, p. 10.

of four is considered a "bare pass," what must an answer contain in order to achieve a score of four and what distinguishes it from the fives and sixes:

It is, of course, possible to make do with the author's analysis or the graders' consensus analysis and to annotate it appropriately to remind the graders of the relative weights and the anomalies and how to treat them. A better approach, however, is to prepare a separate grading guideline that sets forth the issues, the agreed-upon weights assigned to each segment of the answer, directions to the graders on how to treat the expectable permutations, and what constitutes "passing" or better on each of the issues.

The grading guideline needn't be elaborate. On pages 35-34 is a sample of a one that was used in an MBE grading workshop following the July 1995 Multistate Essay Examination. The essay question that underlies it appears at right. In the two-column form, the guideline breaks the question down into three major issues and, under each of the major headings, it sets forth in the left-hand column the sub-points that examinees would be expected to cover in discussing the issue described in the heading.

In keeping with the holistic approach, the guideline states in the right-hand column the gross weights assigned to the major issues. It does not allocate the gross weights further as among the sub-points under each heading. It is left to the grader, for example, to decide how much of the 40% assigned to the first major issue each examinee should receive.

Also in the right-hand column, the guideline contains comments regarding such things as what constitutes a "bare pass," what distinguishes better answers and how expectable examinee responses should be treated. For example, the third call of the essay question at right directs the examinees to discuss the rights of a party under the UCC. In several of the familiarization papers, it was noticed that the examinees were discussing common law contracts and torts principles. It was determined among

Question 2

Acme needed money to finance its manufacturing operations. Brenda agreed to lend Acme \$100,000 if Acme would grant Brenda a security interest in Acme's primary production machine. At that time, the machine was unencumbered by any other security interests. After agreeing to Brenda's terms, Acme delivered to Brenda a properly executed \$100,000 negotiable note payable to the order of Brenda. Immediately upon disbursing the loan funds to Acme, Brenda filed a properly executed financing statement in the appropriate UCC public filing office. She neglected, however, to obtain a written security agreement from Acme.

Two months later, through no fault of Brenda's, Cathy stole the Acme note from Brenda's safe, forged Brenda's signature on the back of the note, and sold the note to Dan, who took the instrument for value, in good faith, and without notice of the theft. Brenda learned that Dan now held the note. Brenda also heard that last week Acme had borrowed \$50,000 from Edward and that Edward also had taken a security interest in Acme's primary production machine.

After hearing about Edward's loan and security interest, Brenda realized that she had never obtained a signed security agreement from Acme granting her an interest in the machine. Upon discovering this oversight, Brenda got Acme to sign such an agreement. By then, however, Edward had already loaned the \$50,000 to Acme, had received Acme's signature on a security agreement, and had properly filed a financing statement covering Acme's machine.

1. Does Brenda's security interest in Acme's machine take priority over Edward's interest? Explain.
2. As between Dan and Brenda, who has superior rights to the note? Explain.
3. If Brenda's rights are superior to Dan's, on what theories, if any, under the UCC, may Dan recover against Cathy? Explain.

GRADING GUIDELINES

ISSUES	COMMENTS
<p>I. PRIORITY AS BETWEEN THE SECURITY INTERESTS OF BRENDA AND EDWARD</p> <ul style="list-style-type: none">Under Article 9 of the UCC, when there is a priority dispute relating to security interest, the party who first filed or perfected gains priority.Perfection requires both attachment of the security interest and filing of notice in the appropriate government office.Brenda's failure to obtain a signed security agreement prevented attachment and, therefore, there was no perfection.Edward's interest was perfected before Brenda's.However, Brenda filed before Edward.Since Brenda later perfected by obtaining a security agreement, her security interest dates back to the time of filing.Since she filed before Edward perfected or filed, Brenda prevails.	<p>40% WEIGHT</p> <ul style="list-style-type: none">Applicants who note these requirements can achieve a "passing" score. Better grades will go to applicants who, in addition, discuss the components of attachment — i.e., (1) the creditor gave value, (2) the debtor had rights in the collateral, and, (3) debtor signed a security agreement.
<p>II. RIGHTS TO THE NOTE AS BETWEEN DAN AND BRENDA</p> <ul style="list-style-type: none">The facts make it appear that Dan was a holder in due course — i.e., that<ul style="list-style-type: none">He gave value;Took in good faith;And without notice of forgery.	<p>40% WEIGHT</p> <ul style="list-style-type: none">We expect better papers to explain why Dan was not an HDC, but applicants who go directly to the "holder" issue and explain why Dan was not a "holder" can achieve a "clear pass."

A MODEL FOR GRADING BAR EXAMINATION ESSAY QUESTIONS 33

GRADING GUIDELINES	
ISSUES	COMMENTS
<ul style="list-style-type: none">One component of being an HDC, however, is that the person must, in the first instance, be a "holder."Since the note was made to the order of Brenda, her indorsement was a necessary component to proper negotiation.Without Brenda's indorsement, Dan could not be a "holder" and, therefore, could not be an HDC.Dan's rights are therefore subject to Brenda's ownership rights.	<ul style="list-style-type: none">Applicants who miss the "holder" issue and, instead, go off on a "real defense" analysis (i.e., that forgery is a "real defense") should receive some credit for reaching the correct result, but cannot get full credit. The call asks who has superior rights to the note, not whether Brenda has a defense against enforcement of the note.
<p>III. DAN'S THEORIES OF RECOVERY AGAINST CATHY</p> <ul style="list-style-type: none">As transferor of the note, Cathy is held to a number of transfer warranties, at least 3 of which she breached:<ul style="list-style-type: none">That she was a person entitled to enforce the note;That all signatures were authentic or authorized;That no defense of any party to the note is good against her.The forgery vitiates all three warranties and gives Dan the right to recover the value paid plus associated damages.	<p>20% WEIGHT</p> <ul style="list-style-type: none">The call specifically asks for Dan's rights under the U.C.C. Thus, discussions about contract and tort (e.g., fraud) theories are not entitled to credit.We expect applicants to discuss the transfer warranties separately. Those who merely note Cathy breached transfer warranties and do not expand on the discussion may achieve a "bare pass," but only if they demonstrate that they understand the law.

the graders that such discussions would not be given credit because they were not responsive to the call of the question, and the guideline, in the comments column, was annotated to reflect this decision.

This kind of grading guideline is also invaluable when a relative scoring system is in place. In order to determine whether one answer is better than another, graders have to begin by agreeing on what they would expect from a "good answer." The description of a "bare pass" on an absolute scale is simply translated to "average" or "just above average" on a relative scale, depending on how well applicants in general are doing in response to the question.

CALIBRATION

Successful implementation of a holistic grading protocol requires training the graders to apply the same standards to all the answers. The term commonly used to describe this training process is "calibration." If only one person is grading all the answers to a particular essay question, there is, of course, no need to coordinate grading standards with anyone else. However, it is still important that the single reader go through the steps of writing an analysis, reading a random sample of answers, adjusting the analysis to account for anomalies, and identifying answers that are examples of each point on the grading scale, sometimes called "anchor papers" or "benchmark answers." Once regular grading is in progress, the reader should refer to the anchor papers periodically in order to maintain a consistent grading standard.⁶

When more than one grader is assigned to the question, it is essential to ensure as much as possible consistency among the graders. This is especially so as the number of graders grows beyond two, because the opportunity for scoring discrepancies expands exponentially.

Calibration can be as simple a process as having the graders meet to review and assign scores to a number of

answers before beginning the final grading. In the course of the meeting, the task for the graders would be to agree to the grading standards and develop the grading guidelines. This works reasonably well if the number of graders per question is small.

When more than two graders are scoring essay or performance test answers, adequate calibration will be achieved if procedures like the ones that follow are employed before the final grading begins.

Initial Meeting — First Phase

After having independently researched and written a narrative analysis to the question, having reviewed each other's written analyses and having read some familiarization papers, the graders meet to construct a consensus analysis and develop the grading guideline. Typically, this part of the meeting takes half a day and proceeds this way.

The supervising grader—who might be a member of the board of bar examiners or someone such as a senior grader—will have prepared a rough draft of a grading guideline based on the author's analysis and the written analyses submitted by the graders. The first order of business is to discuss the differences in the graders' analyses and go over the grading guideline point by point, modifying it to take into account the views expressed by the graders as the discussion progresses. The group will agree to the tentative weights to be assigned to the major issues and will define what points are required in a minimal acceptable answer. Their agreement remains subject to further modification as they proceed through the next step of grading actual answers.

Once a general consensus has been reached on what is in the question and what is expected of the examinees, photocopies of identical sets of about 15 answers drawn at random from the examinee pool are distributed to the graders. They read the first answer silently and assign it a score. The supervising grader tallies the scores, and a discussion ensues, with the graders explaining why they

gave that paper the scores they did. The discussion goes on until the group reaches a consensus, based on the tentative grading guideline, as to what is the correct score for that paper. Then they read the next answer in the batch, discuss it, reach a consensus, and so forth, until they have read and discussed all the photocopied answers and have identified benchmark answers.

When a relative scoring system is employed, graders begin by reading two or three answers and rank-ordering them. The graders then reach consensus on which of the answers was the better or best of the group, and they assign tentative scores of "average," "superior," and so forth. They then read another pair of answers and determine whether they were better or worse than the first group graded. Scores on the sample papers are adjusted until answers that are samples of each of the points on the grading scale are identified; these samples serve as "anchor papers" for actual grading.

Initially at least, there is almost always divergence among the scores assigned by the several graders. As the group works its way through the batch of papers, the divergence narrows, and it becomes easier to identify the points upon which the graders differ. As those points become apparent, the graders decide how to resolve them, and they make the necessary modifications to the grading guideline.

Almost miraculously, there comes a point in the process when the grading differences begin to disappear and, except for the occasional maverick, the scores of the several graders on the same paper converge. At the end of this group grading session, the final version of the grading guideline is agreed to, and it becomes the standard against which all answers are thereafter scored.

Initial Meeting — Second Phase

In order to test the efficacy of the calibration that has gone on so far, the graders are given photocopies of identical sets of another 10 to 20 answers and are directed

to read and score them independently. They record their grades on a score sheet and turn them in to the supervising grader. This process takes up the rest of the first day of calibration. Within the next few days, the supervisor compiles a matrix from the graders' score sheets and identifies any papers on which the scores differed widely enough to be of concern.

Next Meeting

About a week later, the graders reconvene for about half a day, and the first order of business is to discuss the supervising grader's matrix and the papers that he or she identified as having unacceptable score differences. Those papers are read and discussed again to figure out why the scores differed and to reconcile the differences. By now, the supervising grader has prepared and distributed the grading guideline in final form (see the sample on pages 33-34) and, unless a problem arises, the guideline remains unchanged throughout the grading process.

Once again, photocopies of identical sets of 10 or so randomly drawn papers are distributed to the graders, and the procedure that was followed during the first phase of the initial meeting—of reading, scoring, and discussing them one at a time—is repeated. By now, significant score differences on any given paper are few and far between, and the graders are ready to begin actual grading.

FOLLOW-UP CALIBRATION

In jurisdictions where the span of time from the beginning of final grading to completion is great, scores should be reviewed as they are turned in to see if any grader is "drifting" away from the norm, i.e., submitting scores that are consistently higher or lower than those being assigned by other graders of the same question, or, in the case of relative grading, failing to follow guidelines for the percentage of scores that should fall into each

category.⁷ If a problem arises, the supervising grader communicates with that grader and brings him back into line.

A follow-up calibration meeting of all the graders should be scheduled about half way through the grading period. At the follow-up meeting, the procedure of reading, scoring, and discussing a few randomly selected papers is repeated to ensure that the graders have remained calibrated.

CONCLUSION

There is nothing magic or rigid about the procedure described above. It can be modified or truncated in any number of ways to suit the needs of the examining jurisdiction. For example, ten to fifteen graders from jurisdictions that use the NCBE's Multistate Essay Examination usually attend each MEE grading workshop session, the goal of which is to familiarize graders with the range of answers they are likely to encounter. A random sample of three or more answers from each of the represented states—a total of about 30 answers—is available for review. The workshop facilitator reads the answers the day before the workshop and determines which will be used to demonstrate the variety of answers graders might anticipate and which ones demonstrate particular grading problems. The graders need not be calibrated in the sense of agreeing on the difference between an answer that is a "bare pass" and one that is a "bare fail," because each will return to grade papers in his or her jurisdiction, applying state standards to the process. However, graders leave the session in general agreement about how to rank order the papers they have read and having agreed on what constitutes a "good answer" to the question.

The procedures described above lend themselves well to holistic grading and prescribe steps that can be used to ensure that essay examination graders prepare

themselves fully so that they can grade with fairness, consistency, and objectivity.

ENDNOTES

1. See Klein, "Options for Assigning Essay Scores," this issue, p. 24; Lenel, "Grading the Essay Examination," *The Bar Examiner*, August 1990, p. 16.
2. The Merriam-Webster Collegiate Dictionary, 10th Ed., defines holistic as, "relating to or concerned with wholes or with complete systems rather than with the analysis of, treatment of, or dissection into parts."
3. Klein, *Id.* at 24; Lenel, *Id.* at 17.
4. Lenel, *Id.*
5. One of the fail-safe measures used in the preparation of the NCBE's Multistate Essay Examination is to conduct a pretest of the questions before they are finally selected for inclusion on an actual examination form. Recent successful bar applicants are asked to answer the questions under examination conditions, and their answers are looked at to see if any anomalies appear. Occasionally, one does, and the question is modified or the analysis is annotated to alert the graders to make certain allowances in light of the anomaly.
6. Lenel, *Id.* at 20.
7. For example, a state might use a seven-point scale, where the score of four represents a paper that is average, and where graders have been instructed to assign no more than 20% nor less than 10% of their scores to any of the seven categories. On a five-point scale, graders might be instructed to aim for 20% at each of the points of the scale, with no more than 30% nor less than 10% at any point.



Jean C. Gaskill is a retired partner in the San Francisco firm of Brobeck, Phleger & Harrison, where he practiced labor and employment law. He now serves as a mediator and neutral arbitrator in labor and employment disputes. He is a member of the MEE Drafting Committee and chairs the MPT Drafting Committee. He has served as a member of the California Board of Reappraisers

for more than 20 years and as a facilitator of MEE grading workshops since 1988. Jane Peckman Smith, NCBE Director of Testing, contributed to this article.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.